https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

A Study of Small-scale Landslides Susceptibility Dynamic Assessment in South Chian: Merging SBAS-InSAR and Machine Learning

Zhu liang[1,2,3], Jingxin Hou[1,2,3], Yang Liu[1,2,3,*], Ting Wang[4], Guochao Liu[1,2,3], Chunshuai Si[1,2,3] and Jun Wu[1,2,3]

[1]Guangzhou Urban Planning & Design Survey Research Institute Co., Ltd, Guangzhou, 510060, China
[2]Collaborative Innovation Center for Natural Resources Planning and Marine Technology of Guangzhou, Guangzhou, 510060, China
[3]Guangdong Enterprise Key Laboratory for Urban Sensing, Monitoring and Early Warning, Guangzhou, 510060, China
[4]Guangzhou Meteorological Observatory，Guangzhou，511430

** Correspondence: liuyang@gzpi.com.cn*

Abstract. Shallow landslides, triggered by factors such as heavy rainfall and human engineering activities, are characterized by their sudden occurrence, wide distribution, and small scale, which pose significant threats to human life, especially in southern China. Landslide susceptibility assessment (LSA) is crucial for disaster prevention, mitigation, and land-use planning. Traditional assessment methods, such as field surveys and statistical models, predominantly depend on static geological environment factors, which exhibit inherent limitations in capturing spatiotemporal information on dynamic surface deformation. This deficiency directly leads to a lack of timeliness in hazard assessment, which is particularly important for sudden disasters. In this study, small Baseline Subset (SBAS - InSAR) and machine learning were married to explore a dynamic method in LSA, and the Conghua district in Guangzhou was selected as the study area. First, a dataset consisting of 326 historical landslides and 10 static factors, and 2 dynamic factors in the area was prepared. Then, the dataset was divided into two parts, one for modeling training (80%) and the other for testing (20%). Third, factors and samples were involved in the modeling of Random Forest (RF), Light Gradient Boosting Machine (LightGBM), and Extreme Gradient Boosting (XGBoost). Finally, the performance of these models was validated through the area under the curve (AUC) and compared with the models analyzed by 10 static factors only. The results show that the XGBoost model exhibits the best performance, with an AUC of 0.933, and the models considered dynamic factors all performed better than that of static factors only. This study demonstrates that merging SBAS-InSAR and machine learning can provide a reliable technical approach for dynamic assessment of small-scale LSA, which is of great significance for targeted disaster prevention and mitigation in rural and hilly areas of South China.

Keywords: Landslide susceptibility map; Small-scale; SBAS – InSAR; Machine learning; Dynamic evaluation

1. Introduction

Small-scale shallow landslides are a common type of geological disaster, usually occurring in the upper soil layer or weathered rock layer (Ado et al.,2022; Castelli et al.,2025; Sun et al.,2025). Under the combined influence of natural factors (e.g., extreme rainfall, geological weathering) and human activities (e.g., rural road construction, farmland expansion), small-scale landslides have become one of the most frequent geological disasters in southern China (Zhang et al., 2021; Wang et al., 2021). Unlike large-scale landslides, small-scale landslides have small volumes

https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

(usually 10 - 1000 m³) and shallow sliding depths (1 - 3 m), but they are widely distributed, highly sudden, and difficult to monitor (Cui et al.,2023). In recent years, under the influence of global climate change, extreme rainfall events have occurred frequently, and the intensity and frequency of shallow landslides have increased significantly. With the rapid development of urbanization and infrastructure construction, human engineering activities further increase the risk of shallow landslides.

Traditional LSA methods mainly include field survey methods, remote sensing image interpretation methods, and statistical model methods. Field survey methods are time-consuming, labor-intensive, and have limited spatial coverage, making it difficult to carry out large-area assessments (Kovács et al.,2019). Statistical model methods, such as the Logistic Regression model, clustering analysis, require a large number of historical landslide data and cannot handle the nonlinear relationship between factors (Huang et al.,2024; Nwazelibe et al.,2024). In recent years, machine learning methods have been widely used in LSA due to their strong ability to handle nonlinear and high-dimensional data (Huang et al.,2024; Zhou et al.,2025). However, most of these studies use static influencing factors (e.g., lithology, slope, rainfall) and ignore the dynamic changes in slope stability, such as subtle surface deformation before landslide initiation. Small-scale landslides often occur after a period of accelerated surface deformation (e.g., deformation rate increases from -5 mm/year to -15 mm/year within 1 - 2 months before sliding), which is a key precursor of landslide occurrence (Wang et al.,2024).

Traditional remote sensing methods (e.g., optical remote sensing) can only identify existing landslides but cannot capture subtle surface deformation. Field monitoring methods (e.g., inclinometers, GNSS) can measure surface deformation with high accuracy, but they are limited by high cost and low spatial coverage, making it difficult to apply in large - area small - scale landslide assessment (Ebrahim et al.,2024; Nie et al.,2025; Wang et al.,2024). Therefore, the lack of dynamic deformation information is the main bottleneck restricting the accuracy of small-scale landslide susceptibility assessment. Compared with other remote sensing methods (e.g., PS - InSAR), SBAS - InSAR has better adaptability to vegetated areas (which account for over 60% of the study area in South China) by using small baseline interferometric pairs and extracting deformation information from distributed scatterers (e.g., vegetation, soil) (Hussain et al.,2025; Zhang et al.,2025; Guo et al.,2024).

The study is to explore and compare the performance of merging SBAS-InSAR technology with different machine learning models (RF, LightGBM, and XGBoost) in the assessment of small-scale landslide susceptibility. Conghua dissect, Guangzhou was selected as the study area.

2. Study Area and Data Sources

2.1 Overview of the Study Area

The study area is located in the mountainous area of Guangdong Province, China, with a geographical coordinate range of 109°20′ - 110°00′E and 28°10′ - 28°50′N, covering an area of approximately 2,500 km²(Fig.1). The study area has a subtropical monsoon climate, with an average annual rainfall of 1,200 - 1,600 mm, and the rainfall is mainly concentrated in the flood season (May - September), accounting for more than 70% of the annual rainfall. The terrain in the study area is complex and undulating, with an elevation range of 200 - 1,800 m. The main landforms are mountains, hills, and valleys. The stratum in the study area is mainly composed of Devonian sandstone, shale, Carboniferous limestone, and Quaternary loose deposits. The geological structure is relatively complex, with faults and folds developed. The study area is an area with frequent shallow landslide disasters. In

https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

recent years, many shallow landslides have occurred, mainly triggered by heavy rainfall, house and road construction. The number of landslides recorded in the study area is 361, with the volume ranging from 15 m^3 to 400 m^3(Fig.2). These landslides have caused serious damage to local infrastructure and people's lives and property, so it is urgent to carry out a shallow landslide susceptibility assessment in this area.

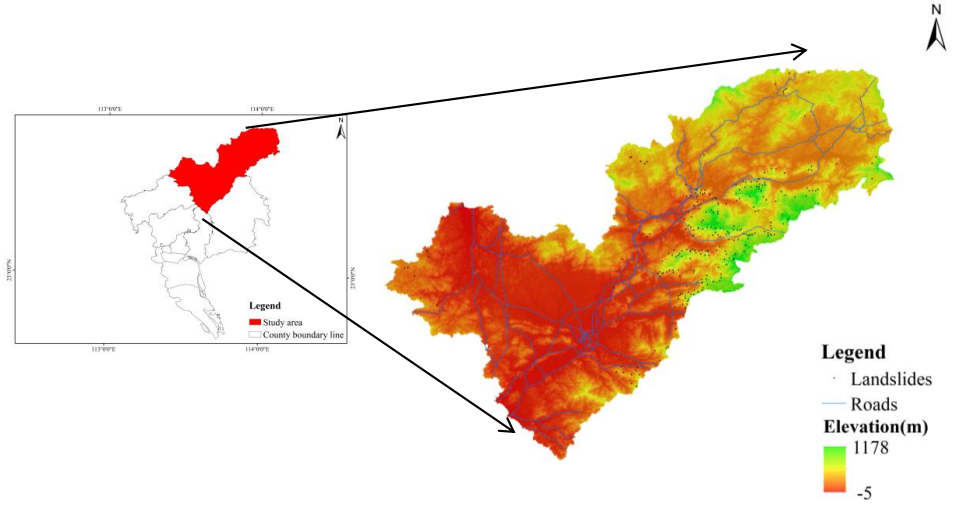


Figure 1: Location of the study area showing elevation and landslide samples.

https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

Figure 2: Photos of the on-site investigation of shallow landslides.

2.2 Data Sources and Preprocessing

2.2.1 Sentinel - 1A/B SAR Data

The SAR data used in this study are Sentinel - 1A/B C - band SAR images provided by the European Space Agency (ESA). The data acquisition period is from January 2024 to December 2025, with a total of 48 images. The imaging mode is Interferometric Wide (IW) mode, with a spatial resolution of 5 m (azimuth) × 20 m (range) and a temporal baseline of 12 days. The main parameters of the Sentinel - 1A/B SAR data are shown in Table 1.

Table 1. Main Parameters of Sentinel - 1A/B SAR Data

Parameter	Value
Satellite	Sentinel - 1A/B
Sensor Band	C - band (5.405 GHz)
Imaging Mode	IW
Spatial Resolution	Azimuth: 5 m; Range: 20 m
Swath Width	250 km
Temporal Baseline	12 days
Polarization	VV + VH
Acquisition Period	January 2024- December 2025
Number of Images	48

The preprocessing of Sentinel - 1A/B SAR data is carried out using the Sentinel Application Platform (SNAP) software.

https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

2.2.2 Digital Elevation Model (DEM) Data

The DEM data used in this study were collected and interpreted independently by the research institution. with a spatial resolution of 10 m. The DEM data are mainly used for geometric correction of SAR images, generation of topographic phase in interferograms, and extraction of topographic factors (slope, aspect, elevation, curvature) for shallow landslide susceptibility assessment.

2.2.3 Geological Data

The geological data used in this study include the 1:50,000 geological map of the study area and the lithology and geological structure reports provided by the local geological survey department. The lithology data are classified into five categories: sandstone, shale, limestone, Quaternary loose deposits, and others. The geological structure data mainly include the distribution of faults and folds. These geological data are digitized using ArcGIS software to form vector data, which are used as important evaluation indices for shallow landslide susceptibility assessment.

2.2.4 Hydrological Data

The hydrological data used in this study include the river distribution data. The river distribution data are obtained from the 1:100,000 topographic map, and the distance from each mapping unit to the nearest river is calculated using ArcGIS software. These hydrological data are used to reflect the influence of rainfall and surface water on shallow landslide occurrence.

2.2.5 Human Activity Data

The human activity data used in this study include the road distribution data. The road distribution data are obtained from the 1:100,000 topographic map, and the distance from each mapping unit to the nearest road is calculated using ArcGIS software. These human activity data are used to reflect the influence of human engineering activities on shallow landslide occurrence.

2.2.6 Normalized Differential Vegetation Index (NDVI)

The vegetation coverage used in this study includes NDVI, which was obtained from the MODIS satellite images provided by NASA.

Table 2. Summary of data information and functions

Items	Scale	Function	Source
Sentinel - 1A/B SAR Data	5 m× 20 m	Extracting surface deformation	European Space Agency (ESA)
DEM	10 m× 10 m	Extraction of topographic factors	Interpreted independently
Geological map	1:50,000	Extraction of lithology and faults	Geological Survey Department
Topographic map	1:100,000	Extraction of river and road distribution	Geological Survey Department
NDVI		Reflecting	NASA

https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

Landslide samples	15 m³ to 400 m³	Vegetation coverage Model training and validation	Survey reports, field surveys, and remote sensing image interpretation

3. Methods

The research framework of this study is shown in Fig.3, which mainly includes 4 parts: SBAS - InSAR surface deformation extraction, data preparation, machine learning modeling, verification, and comparison.

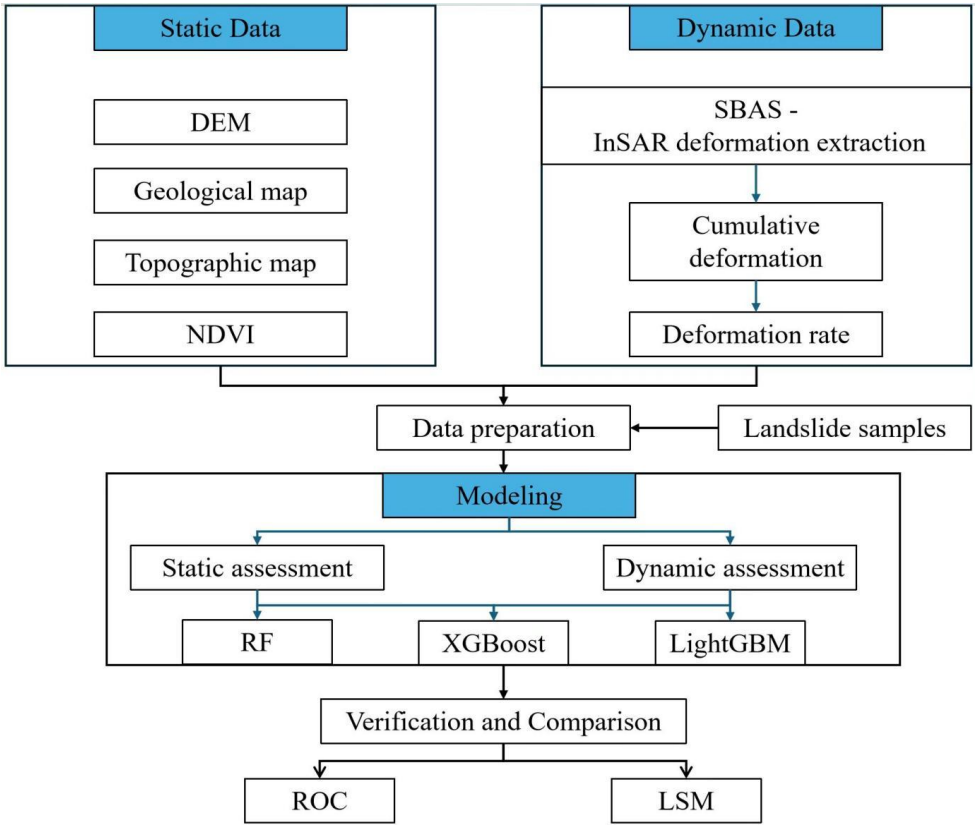


Figure 3: Workflow of the study.

3.1 SBAS - InSAR Method for Surface Deformation Extraction

The Small Baseline Subset (SBAS - InSAR) method is a Time-Series InSAR technique proposed by Berardino et al. (2002). This method selects SAR image pairs characterized by small spatial and temporal baselines, so as to form interferometric pairs. This approach can effectively mitigate the effects of temporal and spatial decorrelation as well as atmospheric delay, thereby enhancing the accuracy of surface deformation measurement:

https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

$$\frac{N+1}{2} \le M \le \frac{N(N+1)}{2} \quad (1)$$

Where N is SAR images, M is interferograms.

Taking t_0 as the reference moment, $\varphi(t_i)$ denotes the differential phase at time t_i (where i = 1, 2, ..., N). For the m-th differential interferogram generated between time instants t_a and t_β (with $t_a > t_\beta$), the interference phase value corresponding to any pixel (x, r) is:

$$\delta\varphi_m(x,r) = \varphi(t_A,x,r) - \varphi(t_B,x,r) \approx \frac{4\pi}{\lambda}[d(t_A,x,r) - d(t_B,x,r)] \quad (2)$$

Where $\delta\varphi_m(x, r)$ denotes the interferometric phase of pixel (x, r) in the m-th interferogram; λ represents the wavelength of the satellite-borne radar; and $d(t_a, x, r)$ as well as $d(t_\beta, x, r)$ respectively stand for the cumulative line-of-sight (LOS) deformations at times t_a and t_β, with both deformations referenced to the reference time t_0.

In accordance with the fundamental principles of the Small Baseline Subset Interferometric Synthetic Aperture Radar (SBAS-InSAR) technique, its standard workflow comprises six core stages, which are specifically defined as follows: data preprocessing, baseline network construction, interferogram generation, phase unwrapping, surface deformation estimation, and geocoding:

$$vm = \frac{\varphi_m - \varphi_{m-1}}{t_m - t_{m-1}} \quad (3)$$

Assuming that the deformation rate between two adjacent interferograms is denoted as v_k, k and (where k and k−1 represent the serial numbers of SAR images, respectively). the value of the *m-th* differential interferogram can be reformulated as follows:

$$\delta\varphi_m = \sum_{k=t,m+1}^{t_A,m}(t_k - t_{k-1})\, v_k \quad (4)$$

In other words, this value represents the integral of the deformation rate v_k over each time period corresponding to the time intervals between the master and slave SAR images. On this basis, Equation (4) can be rewritten into the following matrix form:

$$Bv = \delta\varphi \quad (5)$$

Equation (5) corresponds to a matrix (denoted as B) with a dimension of M×N. When there are multiple small baseline sets involved, the matrix B exhibits rank deficiency. To address this issue, the Singular Value Decomposition (SVD) method can be employed to decompose the coefficient matrix *B*—a process that facilitates the joint inversion of multiple small baseline sets. Accordingly, a least squares solution that achieves the minimization of the norm associated with the cumulative deformation variable can be derived. By virtue of this solution, the accurate estimation of the deformation variable is rendered feasible.

3.2 Data preparation

3.2.1 Landslide Inventory

The historical landslide data used in this study are mainly obtained from the local geological disaster survey reports, field surveys, and remote sensing image interpretation. A total of 361 historical shallow landslide points were collected (as shown in Fig.1), including the location, scale, and occurrence time of the landslides. These samples are used for model training (70%) and validation (30%).

3.2.2 Conditioning factors and mapping unit

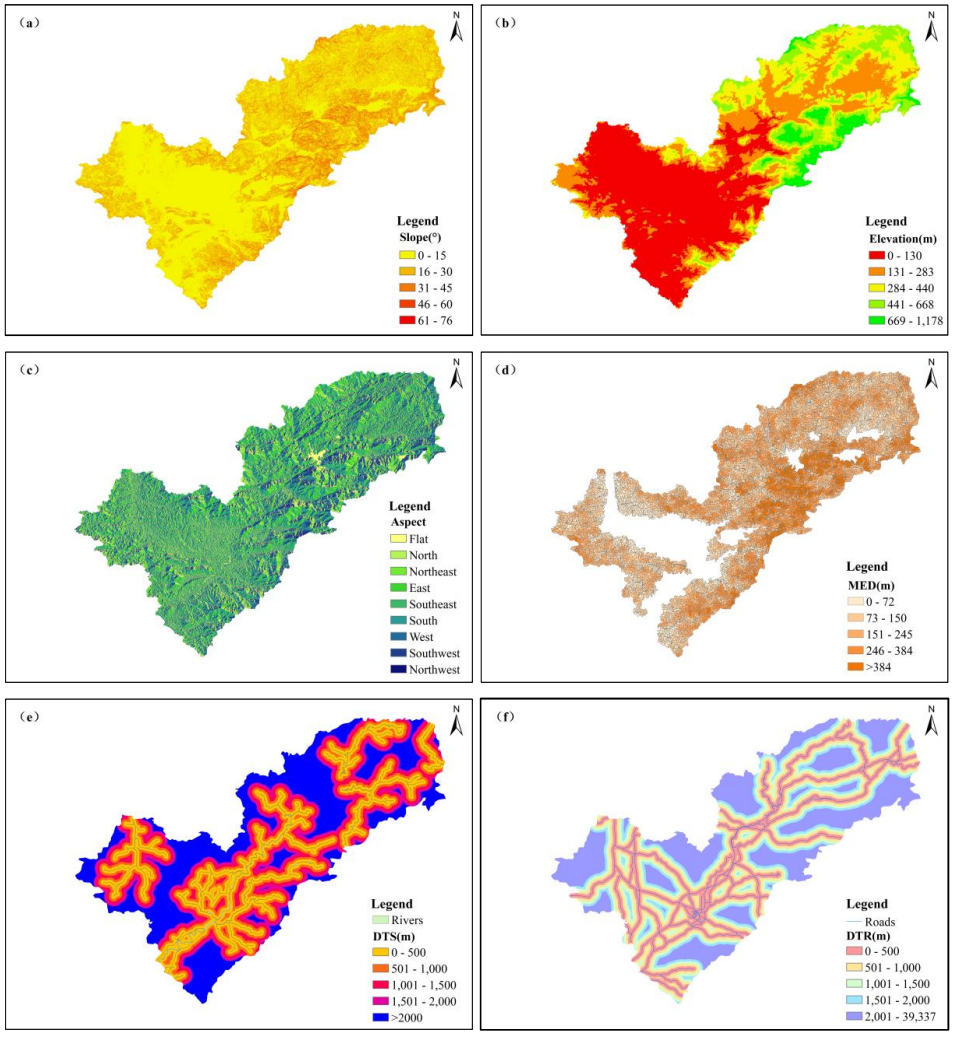
The conditioning factors selected in the study consisted of static and dynamic factors. Static factors included slope,

https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

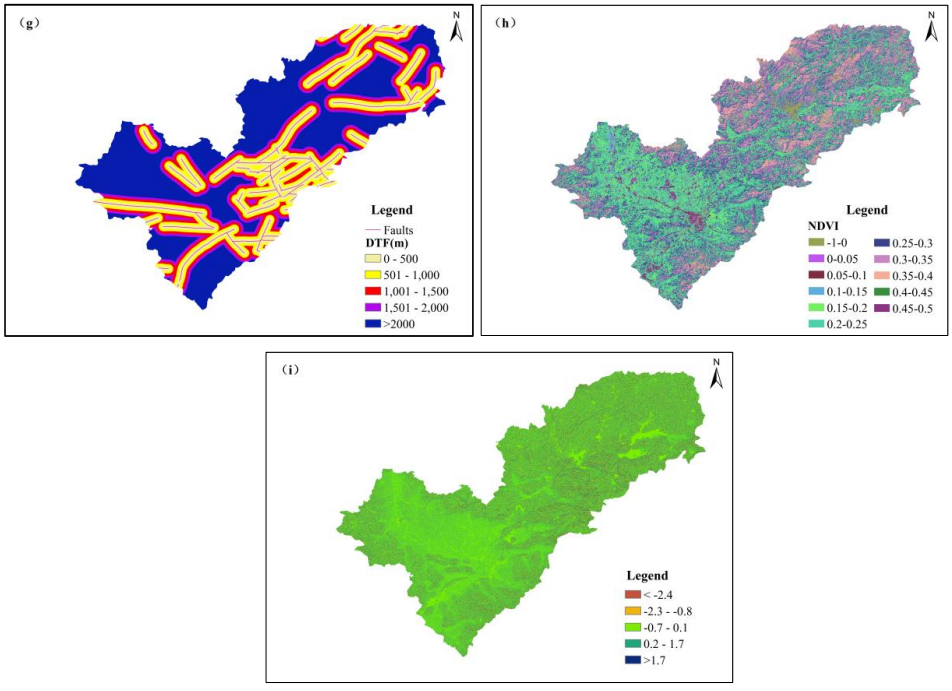
aspect, elevation, curvature, distance to faults, distance to rivers, distance to roads, and NDVI (Fig.4). Dynamic factors included cumulative deformation and deformation rate (Fig.5). The specific indexes and their classification standards are shown in Table 3.

Prior to LSA, a suitable mapping unit must first be determined. Grid cells and slope units are two of the most widely used mapping units, and a detailed comparison of the two can be found in other relevant literature (Lie et al.,2024; Liu et al.,2024). In terms of preserving topographic conditions and distinguishing landslide locations, slope units exhibit superior performance. Accordingly, slope units were adopted as the mapping unit, and the study area was divided into 10426 units.

The study utilized ArcGIS 10.7 for the processing of thematic map generation, unit division, and data analysis, including reclassification, distance analysis, and regional analysis etc.

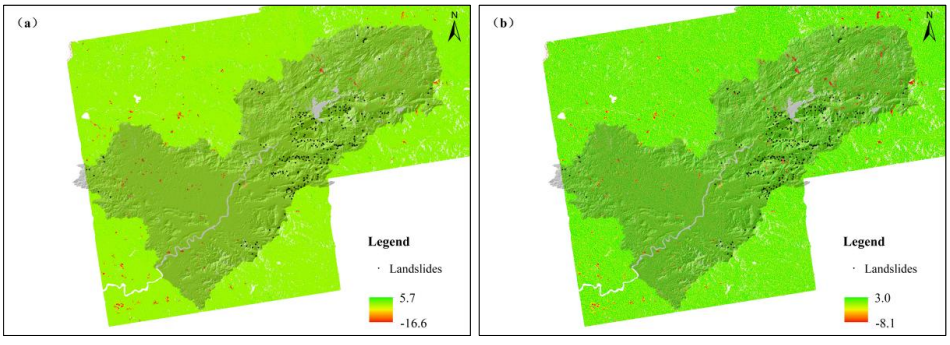


https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026



178
179

180 **Figure 4: Thematic map of static factors:(a) Slope; (b) Elevation; (c) Aspect; (d) Maximum elevation difference; (e)**
181 **distance to rivers; (f) distance to roads; (g) distance to faults; (h) NDVI; (i) Profile curvature.**



182

183 **Figure 5: Thematic map of dynamic factors: (a) Deformation; (b) Deformation rate.**

184 **Table 3. Shallow landslide susceptibility conditioning factors and classification standards**

Category	Conditioning factors	Zone
Static factors	Slope (°)	0~15
		15~30
		30~45
		45~60
		>60
	Aspect	Flat
		North

https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

		East
		Northeast
		Southeast
		South
		Southwest
		West
	Elevation(m)	0~228.4
		228.5~465.8
		465.9~703.2
		703.3~940.6
		>940.6
	Distance to faults(m)	0~500
		500~1000
		1000~1500
		1500~2000
		>2000
	Distance to rivers(m)	0~500
		500~1000
		1000~1500
		1500~2000
		>2000
	Distance to roads(m)	0~500
		500~1000
		1000~1500
		1500~2000
		>2000
	Profile curvature	<-2.4
		-2.3~-0.8
		-0.7~1
		0.1~1.7
		>1.7
	Maximum elevation difference(m)	0~157.8
		157.9~315.6
		315.7~473.4
		473.5~631.2
		>631.3
	NDVI	-1~0.5
Dynamic factors	Cumulative deformation (cm)	-16.6~5.7

https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

	Deformation rate (cm/y)	-8.1~3.0

3.3 Modeling

3.3.1 RF

The Random Forest (RF) algorithm, an ensemble learning method composed of multiple tree-based classifiers, has gained widespread recognition and application in LSA (Li et al.,2024) . During the model training phase, K sample subsets are generated from the original training dataset through the bootstrap sampling technique. The samples that are not included in these subsets, known as out - of - bag (OOB) data, serve as a valuable resource for internal error assessment of the model. Each of these K sample subsets is used to train an individual decision tree, which acts as a base learner, resulting in K distinct classification outcomes (Breiman et al.,2021). For the final classification result, a voting mechanism is adopted instead of a simple weighting approach—each decision tree's classification result is treated as a vote, and the category that receives the majority of votes is determined as the ultimate prediction of the RF model. A key advantage of the RF algorithm is its inherent ability to mitigate over-fitting.

3.3.2 LGBM

The Light Gradient Boosting Machine (LGBM) stands out as a high-efficiency, tree-structured gradient boosting algorithm (Ke et al.,2017). Distinguishing itself from conventional approaches, it adopts a leaf-wise tree growth mechanism with restricted depth, replacing the traditional level-wise strategy—an innovation that simultaneously mitigates overfitting and enhances predictive accuracy. The core of the leaf-wise method lies in prioritizing the leaf node with the maximum gain for splitting at each iteration. In comparison to the level-wise strategy, which expands all nodes at a given depth uniformly, the leaf-wise approach minimizes more of the errors associated with level-wise growth when the total number of splits is identical. Additionally, it reduces redundant computations involving leaf nodes with low gain, thereby guaranteeing the model's computational efficiency without compromising its accuracy.

3.3.3 XGBoost

XGBoost, a robust machine learning framework designed to enhance the scalability of tree-based boosting algorithms, has emerged as a leading and widely adopted method in recent years across various research and application domains (Chen et al.,2015; Zhang e et al.,2023). By leveraging gradient descent optimization to minimize the predefined loss function, the algorithm constructs a predictive model through an ensemble of weak classification trees within a boosting framework. The core idea lies in fitting well the residual of the previous prediction by learning a new function in each iteration; on this basis, it further calculates the score corresponding to each node in accordance with the sample characteristics.

In this study, three machine learning algorithms were employed to model the two scenarios (static factor only and dynamic-static factors coupling) separately, resulting in a total of six models.

3.4 Verification and comparison

The ROC curve is a conventional method for validating the outcomes of LSA. The area under the ROC curve

https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

(AUC, ranging from 0~1) is used to quantify the predictive accuracy of the assessment (Hodasová et al.,2021; Liang et al.,2021; Liang et al.,2023).

The landslide susceptibility map（LSM） was reclassified into five levels as very low(0-0.2), Low(0.2-0.4), moderate(0.4-0.6), high(0.6-0.8), very high (0.8-1). To verify the reliability of the susceptibility map, historical landslide points and the size distribution of areas for each susceptibility level were applied (Liang et al.,2021).

The AUC, sensitivity, specificity, and LSM were applied to verify and compare the effectiveness of different models in the study.

4. Results

4.1 Surface Deformation Results from SBAS - InSAR

Using the SBAS - InSAR method, we obtained the average annual surface deformation rate map of the study area from January 2024 to December 2025. The results show that the surface deformation in the study area is spatially heterogeneous, with deformation rates ranging from -8.1cm/year (subsidence) to 3 cm/year (uplift). Most areas have stable surface conditions (deformation rate < 2cm/year, the green area in Fig.6), accounting for approximately 75% of the total study area. The moderately unstable areas (deformation rate: -1 to -2 cm/year, the yellow area in Fig.6) account for 20%, while the highly unstable areas (-2 to -8 cm/year, the red area in Fig.4) account for 5%.

The highly unstable areas are mainly distributed in the following locations:

1.I t mainly focuses on the areas of Lvian Town, the northern part of Liangkou Town, and the northern part of Wenquan Town.

2. Along the roads, where road construction has disturbed the slope stability, leading to obvious subsidence.

3. Near the fault zones in the eastern part of the study area, where the geological structure is unstable, leading to continuous subsidence.

These highly unstable areas are consistent with the field investigation (Fig.7), indicating that the InSAR-derived surface deformation data can effectively identify potential shallow landslide risk areas. Besides, the majority of these potential geohazard risks are closely associated with human engineering activities—including road construction, residential development, and industrial facility construction—and such hazards typically exhibit a small spatial scale.

https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

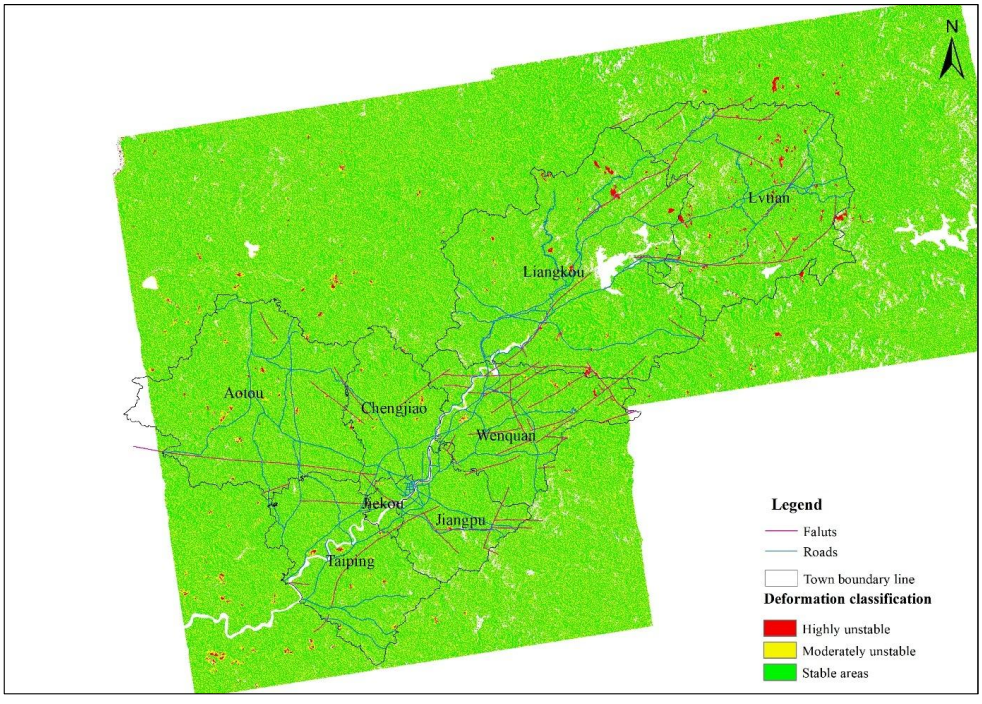


Figure 6: Surface Deformation Rate Map of the Study Area.

Figure 7: Field investigation:(a) slope collapse on 2025.6.17;(b) slope collapse on 2025.6.18 ;(c) landslide on 2025.8.20;(d) landslide on 2025.8.20.

https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

4.2 Model's Performance

The performance of the models was evaluated using confusion matrices, with the detailed results presented in Table 4~5 and Fig.8. For the training dataset, the SAR+XGBoost model achieved the highest sensitivity (0.908), specificity (0.903), accuracy (0.906), and AUC (0.955), followed by the SAR+LGBM model with the values of 0.887, 0.879, 0.883, and 0.943. The RF model performed the worst with the values of 0.834, 0.820, 0.827, and 0.897. While the performance of the LGBM model is average, with the values of 0.842, 0.843, 0.843, and 0.915. The AUC values of all models were all close to 0.9. Among them, the SAR + XGBoost model achieved 0.955, which indicates that the model has achieved an excellent effect in predicting landslide susceptibility. After the integration based on SAR technology, the overall prediction accuracy of the model has improved to a certain extent.

In terms of validation, the SAR+XGBoost model also outperformed the others, reaching 0.865, 0.868, 0.867, and 0.933, whereas the XGBoost obtained 0.808, 0.799, 0.803, and 0.882. Except for the SAR + XGBoost model, the AUC values of all other models are all below 0.9. The RF model also performed the worst with the values of 0.751, 0.736, 0.743, and 0.791. While the performance of the LGBM model is average, with the values of 0.781, 0.77, 0.776, and 0.841. Compared with the training set, the performance of all models has declined to varying degrees, and the AUC value of the RF model has dropped below 0.8 but improved to 0.821 after merging SAR technology. After the integration based on SAR technology, the overall prediction accuracy of the model has improved to a certain extent.

Compared to their performance on the training dataset, all models demonstrated varying degrees of performance degradation on the validation dataset. Specifically, the AUC value of the RF model dropped below 0.8; however, after integrating SAR technology, its AUC value improved to 0.821.

Table 4. Models' performance in training

Indexs	XGBoost	LGBM	RF	SAR+XGBoost	SAR+LGBM	SAR+RF
TP	0.869	0.843	0.817	0.903	0.878	0.864
TN	0.849	0.842	0.837	0.909	0.889	0.878
FP	0.131	0.157	0.183	0.097	0.122	0.136
FN	0.151	0.158	0.163	0.091	0.111	0.122
Sensitivity	0.852	0.842	0.834	0.908	0.887	0.876
Specificity	0.866	0.843	0.820	0.903	0.879	0.866
Accuracy	0.859	0.843	0.827	0.906	0.883	0.871
AUC	0.934	0.915	0.897	0.955	0.943	0.932

Table 5. Models' performance in validation

Indexs	XGBoost	LGBM	RF	SAR+XGBoost	SAR+LGBM	SAR+RF
TP	0.797	0.766	0.727	0.869	0.794	0.766
TN	0.810	0.786	0.759	0.864	0.807	0.786
FP	0.203	0.234	0.273	0.131	0.206	0.234
FN	0.190	0.214	0.241	0.136	0.193	0.214
Sensitivity	0.808	0.781	0.751	0.865	0.804	0.781
Specificity	0.799	0.770	0.736	0.868	0.796	0.770
Accuracy	0.803	0.776	0.743	0.867	0.800	0.776
ROC	0.882	0.841	0.791	0.933	0.866	0.821

https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

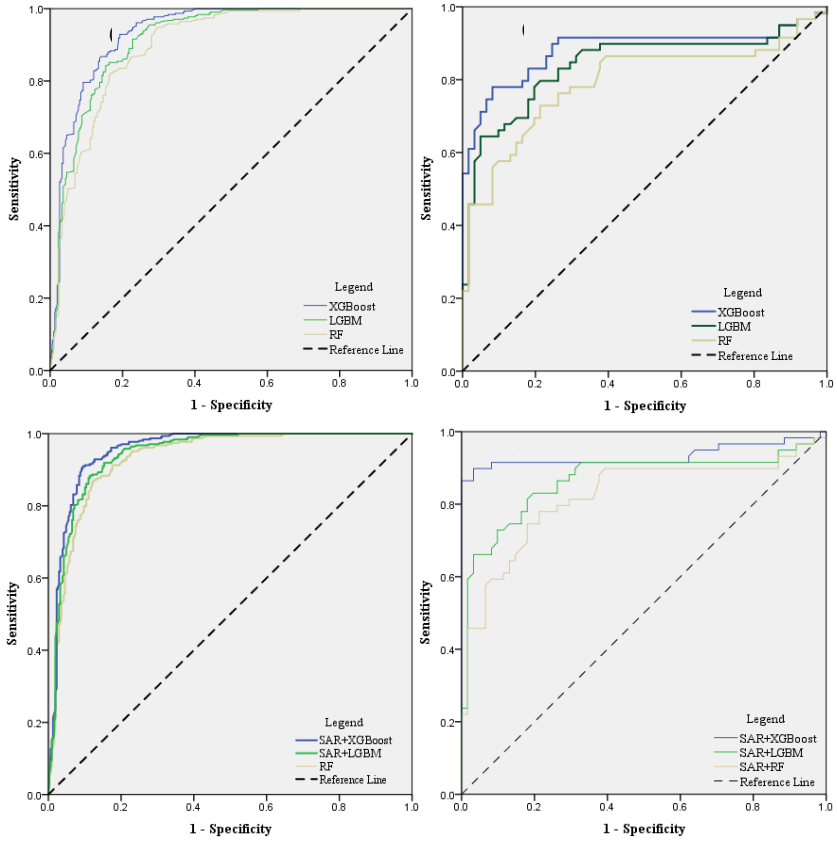


Figure 8: Analysis of ROC curve: (a) traditional machine learning models in the training dataset; (b) traditional machine learning models in the validation dataset; (c) SAR merged models in the training dataset; (d) SAR merged models in the validation dataset.

4.3 Landslide Susceptibility Map

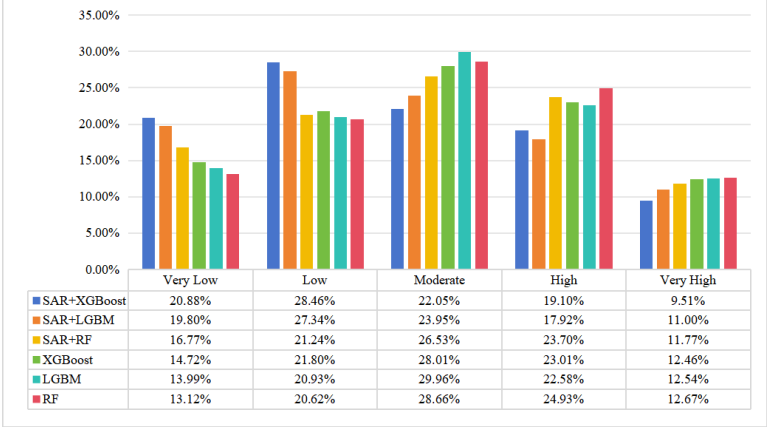
After model training and validation, the landslide susceptibility index (LSI) of each slope unit was calculated, and the landslide susceptibility map of the whole study area was generated. The statistical results of the susceptibility levels are shown in Fig.9. A logically sound landslide susceptibility map must adhere to two fundamental criteria: (1) the density of landslide samples should exhibit a positive correlation with increasing susceptibility classes, with the majority of landslide samples concentrated in the highest susceptibility zone; (2) the map should display spatial continuity and smoothness, with the very high-susceptibility class occupying a relatively small proportion of the total study area (Ado et al.,2022).

The traditional machine learning models consistently predict a relatively high proportion of moderately susceptible areas, with all models yielding values exceeding 28%. In particular, the LGBM model achieved a rate of approximately 30%. For the SAR+XGBoost model, the distribution of predicted susceptibility classes was: 20.88% (very low), 28.46% (low), 22.05% (moderate), 19.10% (high), and 9.51% (very high). While the XGBoost model occupied 14.72% (very low), 21.80% (low), 28.01% (moderate), 23.01% (high), and 12.46(very high), respectively. The RF model occupied 13.12% (very low), 20.62% (low), 28.66% (moderate), 25.83(high), and

https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

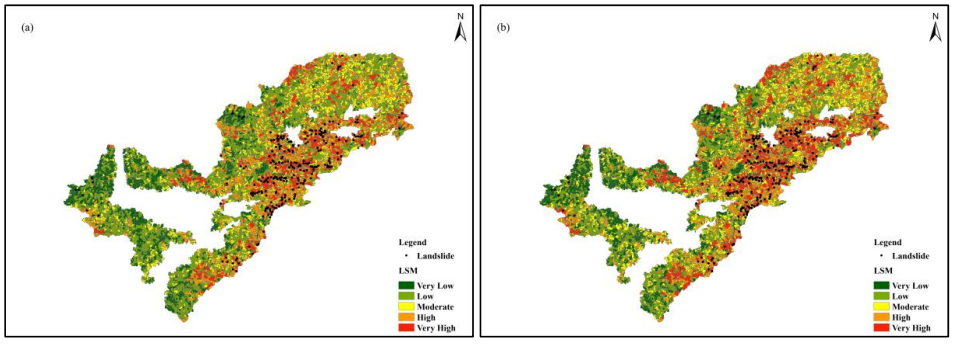
11.77(very high), respectively. While the SAR+RF model occupied 16.77% (very low), 21.24% (low), 26.53% (moderate), 23.70%(high), and 11.77% (very high), respectively.

As illustrated in Fig.10, the landslide samples (represented by dark spots) are predominantly distributed in the red-colored regions (corresponding to high/very high susceptibility), while a few samples are scattered improperly in the green area (corresponding to low/very low susceptibility). Furthermore, the high and very high susceptibility zones are spatially concentrated in the eastern region, which is consistent with the actual spatial distribution of historical landslides. This subregion is characterized by steep slopes (slope>30°) with substantial elevation differences (MED>300), close proximity to the adjacent road network (DTR<500), and a nearby distribution of geological faults. The low and very low susceptibility zones are spatially concentrated in the western region. This region is characterized by relatively gentle terrain with minimal elevation variations and a sparse distribution of faults. Collectively, these observations demonstrate that the susceptibility maps generated by the 6 models are generally logically consistent and reliable, but there are certain differences. Furthermore, as illustrated in Fig.11, the high-deformation zones identified via SAR remote sensing do not fully overlap with the high-susceptibility areas delineated by the model; correspondingly, the stable zones detected by SAR technology also exhibit incomplete spatial overlap with the low-susceptibility areas. The highly unstable area is located in the plain area of the residential zone.



	Very Low	Low	Moderate	High	Very High
SAR+XGBoost	20.88%	28.46%	22.05%	19.10%	9.51%
SAR+LGBM	19.80%	27.34%	23.95%	17.92%	11.00%
SAR+RF	16.77%	21.24%	26.53%	23.70%	11.77%
XGBoost	14.72%	21.80%	28.01%	23.01%	12.46%
LGBM	13.99%	20.93%	29.96%	22.58%	12.54%
RF	13.12%	20.62%	28.66%	24.93%	12.67%

Figure 9: Percentages of areas in LSM.



https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

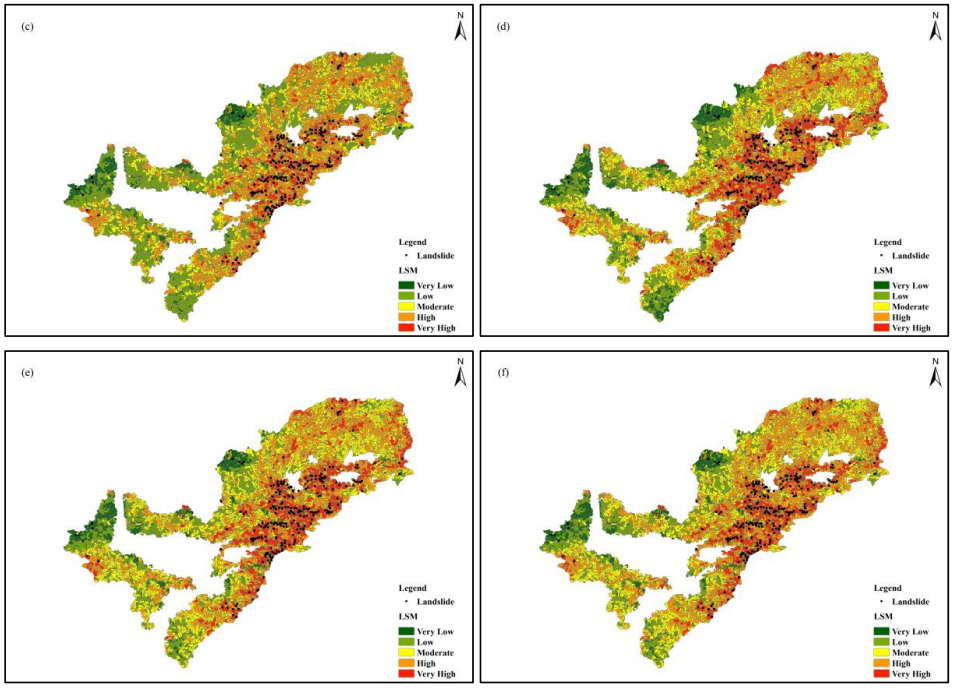
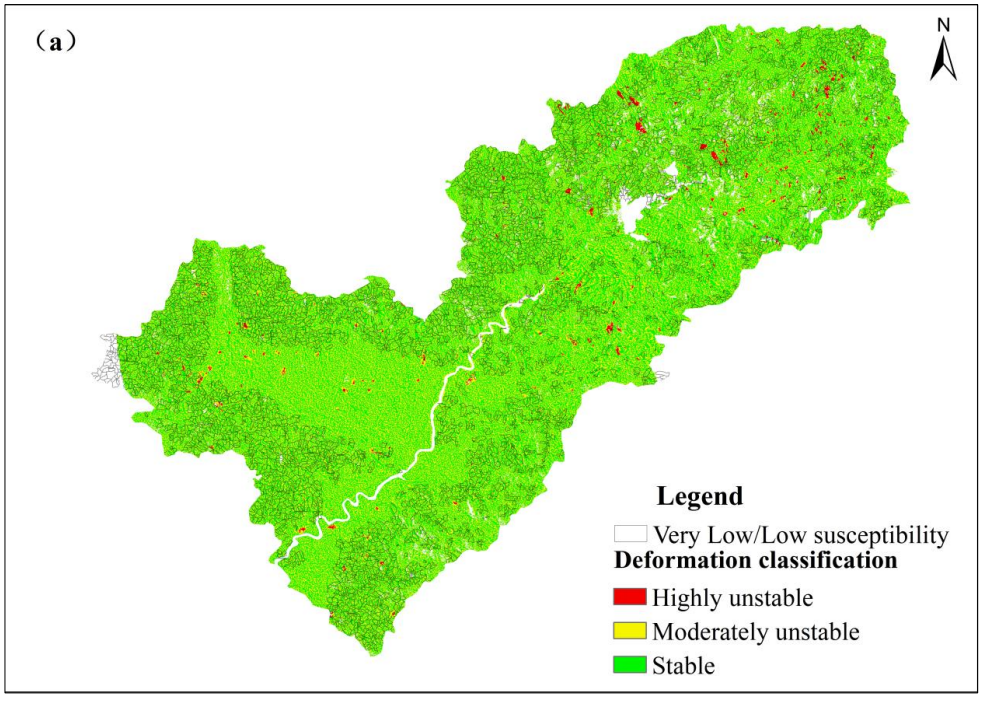


Figure 10: LSM produced by: (a) SAR+XGBoost; (b) SAR+LGBM; (c) SAR+RF; (d) XGBoost; (e) LGBM; (f) RF.



https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

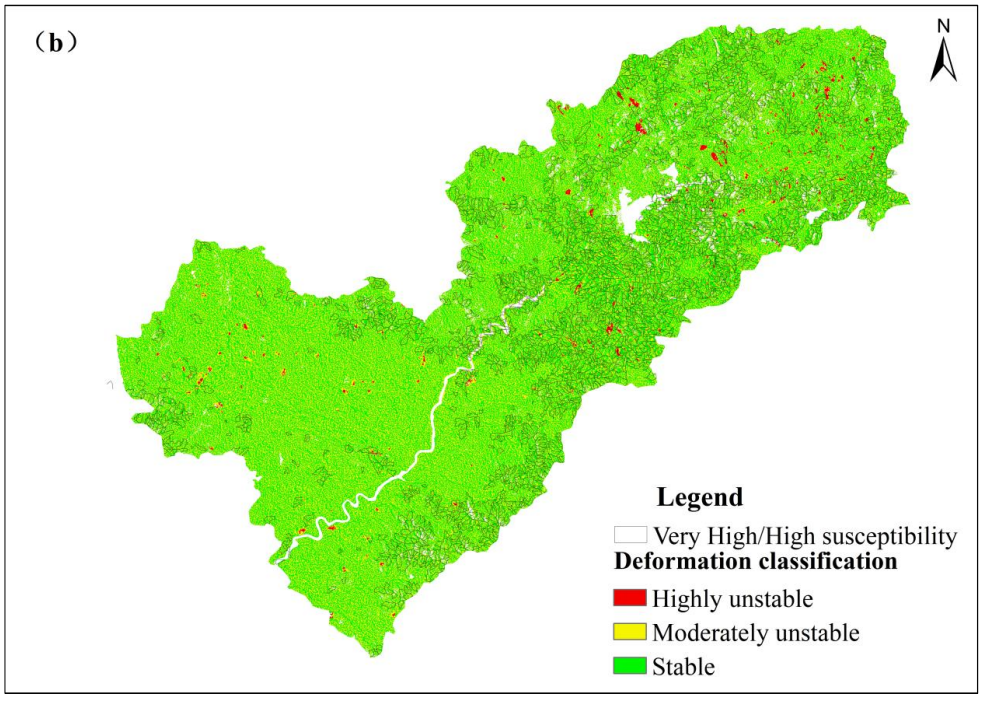
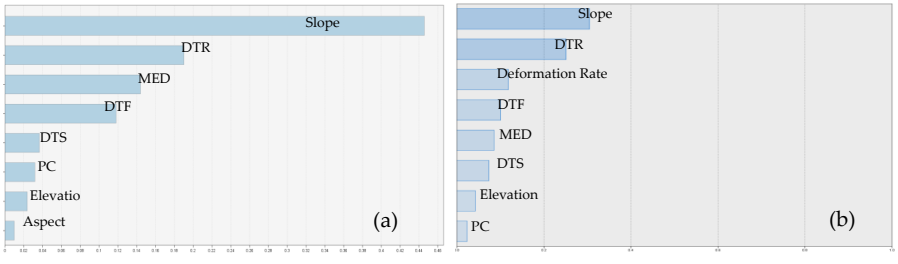


Figure 11: LSM produced by SAR+XGBoost: (a) Low/very low susceptibility; (b) High/very high susceptibility.

4.4 Analysis of the Importance of Conditioning Factors

To quantify the critical contribution of dynamic factors to landslide susceptibility prediction, the Gini index—an indicator that measures the correlation between predictive variables and model outputs—was employed in this study, and the results were normalized (Dou et al.,2023; Ye et al.,2022). In the static factor analysis, the key contributing variables were identified as the slope, DTR, MED, and DTF, with these factors exhibiting relatively high weight coefficients of 0.45, 0.19, 0.14, and 0.12, respectively (Fig.12; Table 6). By contrast, DTS, PC, elevation, and aspect yielded lower weight coefficients, indicating that these factors exert a relatively minor influence on landslide occurrence.

After coupling with SAR technology, the key contributing variables were identified as the slope, DTR, Deformation rate (DR), and DTF, with these factors exhibiting relatively high weight coefficients of 0.3, 0.25, 0.12, and 0.1, respectively (Fig.12; Table 7). By contrast, MED, DTS, elevation, and PC yielded lower weight coefficients, indicating that these factors exert a relatively minor influence on landslide occurrence.



https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

Figure 12: Factor Importance Analysis: (a) Static factor analysis; (b) Dynamic and static factor coupling analysis.

Table 6 Static Factor Importance Analysis

Factors	Slope	DTR	MED	DTF	DTS	PC	Elevation	Aspect
Gini	0.45	0.19	0.14	0.12	0.04	0.03	0.02	0.01

Table 7 Dynamic and static factor Importance Analysis

Factors	Slope	DTR	DR	DTF	MED	DTS	Elevation	PC
Gini	0.3	0.25	0.12	0.1	0.09	0.07	0.04	0.02

5. Discussion

5.1 Models' comparison

Different algorithms exhibit distinct focuses and design priorities, and their performance typically varies depending on the specific research context, characteristics of the study area, and quality of input samples (Hussin et al.,2016; Lai et al.,2019; Yilmaz et al.,2010). With the advancement of related technologies and the maturation of theoretical frameworks, a variety of advanced algorithms—such as convolutional neural networks (CNNs)—have been developed and increasingly applied to landslide susceptibility mapping (LSM) (Wang et al.,2019; Zhao et al.,2024). These algorithmic advancements aim to enhance the prediction accuracy and reliability of LSM results. The three algorithms adopted in this study are all tree-based ensemble algorithms. Regarding accuracy, the performance of the three algorithms was similar throughout the training and validation stages. Overall, the integration of SAR technology contributed to a measurable improvement in the general prediction accuracy of the models. Tree-based ensemble algorithms are capable of efficiently learning discriminative features from SAR-derived ground deformation interpretation data, illustrating the utility of incorporating this remote sensing technology to enhance the generalization capability of landslide susceptibility assessment models.

For LSM, the six models yield generally consistent predictions of the overall susceptibility spatial trends. However, traditional machine learning methods tend to delineate a larger proportion of the study area as moderately susceptible zones, which increases the uncertainty of the prediction results. Models integrated with SAR technology exhibit a reduced proportion of high/very high landslide susceptibility areas, accompanied by a corresponding increase in the proportion of low and very low susceptibility areas.

5.2 Contribution of InSAR-Derived Deformation Data to Susceptibility Assessment

In this study, ground deformation is treated as a dynamic influencing factor and incorporated into the comprehensive evaluation framework for landslide susceptibility. Compared with models that solely incorporate static factors, the algorithms integrated with SAR technology not only achieve improved prediction accuracy but also yield more rational spatial distribution patterns across different landslide susceptibility classes (Hua et al.,2021; Lamichhane et al.,2025). Through the analysis of the importance of the conditioning factors, deformation rate was essential (Devara et al.,2021).

Most prior studies rely on detailed landslide inventories, operating under the implicit assumption that the spatiotemporal patterns of future landslides will be consistent with historical and contemporary observations (Steger et al.,2017; Huang et al.,2024; Zhi et al.,2023). However, intensive human activities and sustained rainfall

https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

events frequently induce novel landslide patterns, thereby challenging the validity of this fundamental premise. The deployment of landslide monitoring equipment is not only associated with substantial economic costs but also faces challenges in providing effective early warning for sudden-onset geohazards. Notably, in southern China, slope cutting activities associated with rural residential construction and road development are widespread, and geological hazards are characterized by high frequency, extensive spatial distribution, and predominantly small scales, rendering comprehensive coverage by monitoring systems technically infeasible and economically impractical (Casagli et al.,2023).

Static landslide susceptibility assessment models tend to overestimate the stability of slopes that have been engineered for stabilization, while underestimating the instability of slopes disturbed by anthropogenic activities (Singh et al.,2010). Furthermore, the calibration of static landslide susceptibility assessment models must be updated in tandem with the refinement and expansion of landslide inventories. However, the refinement and expansion of landslide inventories are both financially costly and logistically infrequent. Consequently, the absence of timely updated landslide inventories readily leads to the underestimation of emerging landslide patterns triggered by urban expansion or shifts in meteorological event regimes.

Compared with traditional assessment methods that ignore dynamic deformation information, the integration of InSAR data provides two key improvements. 1. **Early Identification of Potential Risks**: Traditional methods mainly rely on static factors (e.g., lithology, slope) to identify high-risk areas, but these factors cannot reflect the dynamic changes in slope stability. The InSAR-derived deformation data can capture subtle subsidence (e.g., -5 to -10 mm/year) that is imperceptible to field surveys, enabling early identification of slopes in the "pre-landslide" stage; 2. **Improvement of Assessment Accuracy**: The validation results show that the overall accuracy of the susceptibility map is 15 - 20% higher than that of traditional static assessment methods. This significant improvement confirms that InSAR data effectively fills the gap of dynamic information in traditional assessments (Chen et al.,2023).

However, the high- and low-deformation zones identified via SAR remote sensing do not precisely correspond to the high- and low-susceptibility areas predicted by the landslide susceptibility model. SBAS - InSAR also has limitations. For example, in areas with extremely high vegetation coverage (e.g., dense primeval forests), the coherence of SAR images is still low, leading to missing deformation data. Conversely, geological hazard occurrences exhibit inherent spatiotemporal randomness, while the current SAR remote sensing systems are constrained by finite spatiotemporal resolution. Therefore, the evaluation of landslide susceptibility should incorporate both static and dynamic factors.

6. Conclusions

A high-quality landslide susceptibility map provides clear guidance for local disaster prevention and mitigation work. This study systematically explored the application of SBAS-InSAR in small-scale landslide susceptibility dynamic assessment, using tree-based ensemble algorithms and the Conghua district as the study area. The main conclusions are as follows:

1. The SBAS-InSAR technique enables the effective capture of spatiotemporal dynamic surface deformation information, which is of considerable significance for the identification and monitoring of small-scale landslide hazards in the southern China region.

2. Following the integration of deformation-derived dynamic factors, LSA based on the tree-based algorithm

https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

has achieved significant improvements in prediction accuracy and reliability—demonstrating the applicability of fusing SAR-derived deformation data with machine learning algorithms for dynamic landslide risk evaluation.

However, there are some limitations of the Study:

1. Diverse SAR technologies (e.g., SBAS-InSAR, PS-InSAR, or TomoSAR) can be leveraged to extract higher-dimensional deformation features, which can be incorporated into LSA frameworks.

2. The integrated technology requires rigorous validation through multiple heterogeneous algorithms and cross-validation across distinct geomorphological study areas with varying geological.

Data availability

The data used to support the findings of this study are included within the article.

Author Contributions

Z.L., writing—original draft, methodology, and software; W.P., W.L. and H.H., review and validation; J.H. and K.L., reviewing and editing. K.J. and G.L. are responsible for the review and validation. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Financial support

The completion of this work was supported by Collaborative Innovation Center for Natural Resources Planning and Marine Technology of Guangzhou (No. 2023B04J0301, No. 2025B04J0031), the Science and Technology Foundation of Guangzhou Urban Planning & Design Survey Research Institute (Grant No. RDI2240204004). Last but not least, many thanks for the support of Academic Specialty Group for Urban Sensing in Chinese Society of Urban Planning.

References

1. Ado M, Amitab K, Maji A K, et al. Landslide susceptibility mapping using machine learning: A literature survey[J]. Remote Sensing, 2022, 14(13): 3029.
2. Castelli M, Filipello A, Fasciano C, et al. Geohazard Plugin: A QGIS Plugin for the Preliminary Analysis of Landslides at Medium–Small Scale[J]. Land, 2025, 14(2): 290.
3. Sun L, Zhao Y, Sun Y, et al. Study on the reactivation mechanism and weak substrate effect of bottom-saturated reservoir old landslides–based on a small-scale model[J]. Bulletin of Engineering Geology and the Environment, 2025, 84(9): 420.
4. Zhang K, Song H, Sun S, et al. Distribution and genesis of submarine landslides in the northeastern South China Sea[J]. Geological Journal, 2021, 56(3): 1187-1201.
5. Wang D, Hao M, Chen S, et al. Assessment of landslide susceptibility and risk factors in China[J]. Natural hazards, 2021, 108(3): 3045-3059.
6. Cui Y, Qian Z, Xu W, et al. A Small-Scale Landslide in 2023, Leshan, China: Basic Characteristics, Kinematic Process and Cause Analysis[J]. Remote Sensing, 2024, 16(17): 3324.
7. Kovács I P, Czigány S, Dobre B, et al. A field survey–based method to characterise landslide development: a case

https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

study at the high bluff of the Danube, south-central Hungary[J]. Landslides, 2019, 16(8): 1567-1581.
8. Huang S, Chen L. Landslide susceptibility mapping using an integration of different statistical models for the 2015 Nepal earthquake in Tibet[J]. Geomatics, Natural Hazards and Risk, 2024, 15(1): 2396908.
9. Nwazelibe V E, Egbueri J C. Geospatial assessment of landslide-prone areas in the southern part of Anambra State, Nigeria using classical statistical models[J]. Environmental Earth Sciences, 2024, 83(7): 220.
10. Huang F, Xiong H, Jiang S H, et al. Modelling landslide susceptibility prediction: A review and construction of semi-supervised imbalanced theory[J]. Earth-Science Reviews, 2024, 250: 104700.
11. Hong H, Wang D, Zhu A X, et al. Landslide susceptibility mapping based on the reliability of landslide and non-landslide sample[J]. Expert Systems with Applications, 2024, 243: 122933.
12. Zhou C, Gan L, Cao Y, et al. Landslide susceptibility assessment of the Wanzhou district: Merging landslide susceptibility modelling (LSM) with InSAR-derived ground deformation map[J]. International Journal of Applied Earth Observation and Geoinformation, 2025, 136: 104365.
13. Wang C, Yang Y, Zhang X, et al. Ultra-High Sensitivity Real-Time Monitoring of Landslide Surface Deformation via Triboelectric Nanogenerator[J]. Advanced Materials, 2024, 36(50): 2410471.
14. Ebrahim K M P, Gomaa S M M H, Zayed T, et al. Recent phenomenal and investigational subsurface landslide monitoring techniques: a mixed review[J]. Remote Sensing, 2024, 16(2): 385.
15. Nie W, Tian C, Song D, et al. Disaster process and multisource information monitoring and warning method for rainfall-triggered landslide: a case study in the southeastern coastal area of China[J]. Natural Hazards, 2025, 121(3): 2535-2564.
16. Wang L, Zhang K, Chen Y, et al. Progressive deformation mechanism of colluvial landslides induced by rainfall: insights from long-term field monitoring and numerical study[J]. Landslides, 2024, 21(12): 3069-3086.
17. Hussain S, Pan B, Hussain W, et al. Integrated PSInSAR and SBAS-InSAR analysis for landslide detection and monitoring[J]. Physics and Chemistry of the Earth, Parts A/B/C, 2025: 103956.
18. Zhang J, Zuo X, Li Y, et al. Detection and assessment of potential landslides in the Xiaojiang River Basin using SBAS-InSAR[J]. Scientific Reports, 2025, 15(1): 16082.
19. Guo J, Xi W, Yang Z, et al. Landslide hazard susceptibility evaluation based on SBAS-InSAR technology and SSA-BP neural network algorithm: A case study of Baihetan Reservoir Area[J]. Journal of Mountain Science, 2024, 21(3): 952-972.
20. Berardino, P., Fornaro, G., Lanari, R., & Sansosti, E. (2002). A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. *IEEE Transactions on Geoscience and Remote Sensing, 40*(11), 2375 - 2383.
21. Lie X U, Nengpan J U, Mingdong D, et al. Comparative analysis of landslide susceptibility assessment accuracy in different evaluation units[J]. Journal of Engineering Geology, 2024, 32(5): 1640-1653.
22. Liu S, Du J, Yin K, et al. Regional early warning model for rainfall induced landslide based on slope unit in Chongqing, China[J]. Engineering Geology, 2024, 333: 107464.
23. Li M, Wang H, Chen J, et al. Assessing landslide susceptibility based on the random forest model and multi-source heterogeneous data[J]. Ecological Indicators, 2024, 158: 111600.
24. Breiman L. Random forests[J]. Machine learning, 2001, 45(1): 5-32.
25. Ke G, Meng Q, Finley T, et al. Lightgbm: A highly efficient gradient boosting decision tree[J]. Advances in neural information processing systems, 2017, 30.
26. Chen T, He T, Benesty M, et al. Xgboost: extreme gradient boosting[J]. R package version 0.4-2, 2015, 1(4): 1-4.
27. Zhang J, Ma X, Zhang J, et al. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model[J]. Journal of environmental management, 2023, 332: 117357.
28. Hodasová K, Bednarik M. Effect of using various weighting methods in a process of landslide susceptibility assessment[J]. Natural Hazards, 2021, 105(1): 481-499.
29. Liang Z, Wang C, Duan Z, et al. A hybrid model consisting of supervised and unsupervised learning for landslide susceptibility mapping[J]. Remote Sensing, 2021, 13(8): 1464.
30. Liang Z, Peng W, Liu W, et al. Exploration and comparison of the effect of conventional and advanced modeling algorithms on landslide susceptibility prediction: A case study from Yadong Country, Tibet[J]. Applied Sciences, 2023, 13(12): 7276.
31. Liang Z, Wang C, Khan K U J. Application and comparison of different ensemble learning machines combining with a novel sampling strategy for shallow landslide susceptibility mapping[J]. Stochastic Environmental Research and Risk Assessment, 2021, 35(6): 1243-1256.
32. Ado M, Amitab K, Maji A K, et al. Landslide susceptibility mapping using machine learning: A literature survey[J]. Remote Sensing, 2022, 14(13): 3029.
33. Dou H, He J, Huang S, et al. Influences of non-landslide sample selection strategies on landslide susceptibility

https://doi.org/10.5194/egusphere-2026-218
Preprint. Discussion started: 23 February 2026

mapping by machine learning[J]. Geomatics, Natural Hazards and Risk, 2023, 14(1): 2285719.
34. Ye C, Wei R, Ge Y, et al. GIS-based spatial prediction of landslide using road factors and random forest for Sichuan-Tibet Highway[J]. Journal of Mountain Science, 2022, 19(2): 461-476.
35. Hussin H Y, Zumpano V, Reichenbach P, et al. Different landslide sampling strategies in a grid-based bi-variate statistical susceptibility model[J]. Geomorphology, 2016, 253: 508-523.
36. Lai J S, Chiang S H, Tsai F. Exploring influence of sampling strategies on event-based landslide susceptibility modeling[J]. ISPRS International Journal of Geo-Information, 2019, 8(9): 397.
37. Yilmaz I. The effect of the sampling strategies on the landslide susceptibility mapping by conditional probability and artificial neural networks[J]. Environmental Earth Sciences, 2010, 60(3): 505-519.
38. Wang Y, Fang Z, Hong H. Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China[J]. Science of the total environment, 2019, 666: 975-993.
39. Zhao Z, Chen T, Dou J, et al. Landslide susceptibility mapping considering landslide local-global features based on CNN and transformer[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024, 17: 7475-7489.
40. Hua Y, Wang X, Li Y, et al. Dynamic development of landslide susceptibility based on slope unit and deep neural networks[J]. Landslides, 2021, 18(1): 281-302.
41. Lamichhane S, Kansakar A R, Devkota N, et al. Integrating dynamic factors for predicting future landslide susceptibility[J]. Environmental Earth Sciences, 2025, 84(3): 89.
42. Devara M, Tiwari A, Dwivedi R. Landslide susceptibility mapping using MT-InSAR and AHP enabled GIS-based multi-criteria decision analysis[J]. Geomatics, Natural Hazards and Risk, 2021, 12(1): 675-693.
43. Steger S, Brenning A, Bell R, et al. The influence of systematically incomplete shallow landslide inventories on statistical susceptibility models and suggestions for improvements[J]. Landslides, 2017, 14(5): 1767-1781.
44. Huang F, Mao D, Jiang S H, et al. Uncertainties in landslide susceptibility prediction modeling: A review on the incompleteness of landslide inventory and its influence rules[J]. Geoscience Frontiers, 2024, 15(6): 101886.
45. Zhiyong F, Changdong L, Wenmin Y. Landslide susceptibility assessment through TrAdaBoost transfer learning models using two landslide inventories[J]. Catena, 2023, 222: 106799.
46. Casagli N, Intrieri E, Tofani V, et al. Landslide detection, monitoring and prediction with remote-sensing techniques[J]. Nature Reviews Earth & Environment, 2023, 4(1): 51-64.
47. Singh T N, Verma A K, Sarkar K. Static and dynamic analysis of a landslide[J]. Geomatics, Natural Hazards and Risk, 2010, 1(4): 323-338.
48. Chen L, Ma P, Yu C, et al. Landslide susceptibility assessment in multiple urban slope settings with a landslide inventory augmented by InSAR techniques[J]. Engineering Geology, 2023, 327: 107342.